

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-284517

(43) 公開日 平成11年(1999)10月15日

(51) IntCl.⁶

H 0 3 M 7/30

識別記号

F I

H 0 3 M 7/30

Z

審査請求 未請求 請求項の数19 OL (全 14 頁)

(21) 出願番号 特願平10-367699

(22) 出願日 平成10年(1998)12月24日

(31) 優先権主張番号 9 9 9 5 1 8

(32) 優先日 1997年12月29日

(33) 優先権主張国 米国 (U S)

(71) 出願人 598177441

ブルース ケイ マーティン

Bruce K. Martin

アメリカ合衆国 カリフォルニア州

94306 バロ・アルト ウェルズベリ・コ

ート 604

(72) 発明者 ブルース ケイ マーティン

アメリカ合衆国 カリフォルニア州

94306 バロ・アルト ウェルズベリ・コ

ート 604

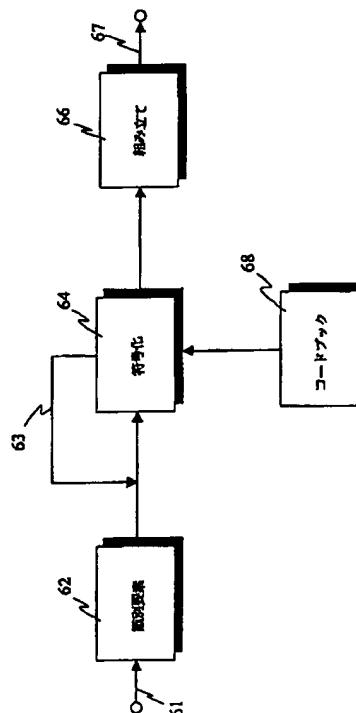
(74) 代理人 弁理士 伊東 忠彦 (外1名)

(54) 【発明の名称】 シンタックス構造を保持してマークアップ言語を用いたドキュメントを圧縮する方法

(57) 【要約】

【課題】 移動電話機とインターネットのようなネットワーク間の通信チャネルは帯域が非常に制限されている。マークアップ言語のような種々の形式で表現されるドキュメントの伝送を、要素のシンタックス的特徴が符号化表示からすぐに決定され得るようなコードヘドドキュメント要素を圧縮することにより、より効率的にする。

【解決手段】 マークアップ言語のタグ属性及び内容のようなシンタックス情報の存在の表示は、予め定められたコードの開始を基準とする位置に伝達される。好ましくは、その位置は表される要素のタイプと独立である。このようにして、ドキュメント要素の圧縮又は符号化表示は、展開又は復号化の必要なしに効率的に処理され得る。加えて、マークアップ言語へなされる将来の拡張は、その新たな拡張を認識しない符号器及び復号器を存在させることにより効率的に処理され得る。



【特許請求の範囲】

【請求項1】 ドキュメントを表す入力情報の容量の必要量を減少させる方法において、

ドキュメントを表す入力情報を受信し、複数の要素を識別し、各要素は各々のタイプを有し、少なくともいくつかの要素は1つ又はそれ以上の各々のシンタックス的特徴を表すシンタックス情報を有し、

複数のコードを生成し、各々のコードは開始を有し、表された部分の情報の必要容量よりも低い必要容量を有する形で少なくとも各々の要素の部分を表し、該各々のコードは、各々の要素タイプと、該各々の要素のシンタックス情報の存在又は非存在を示すシンタックス表示とを伝達し、該各々のコードは該各々のコードの開始を基準とした所定の位置に該シンタックス表示を伝達し、複数のコードと該複数のコードでは表されない複数の要素の部分とを伝送又は格納に適する形に組み立てることによりドキュメントを表す符号化情報を生成することを特徴とする方法。

【請求項2】 前記コードは、要素タイプとは独立に、シンタックス表示がシンタックス情報の存在又は非存在を示す形で生成されることを特徴とする請求項1に記載の方法。

【請求項3】 前記方法は更に、複数のコードブックから1つのコードブックを選択することからなり、コードのうち少なくともいくつかは該選択されたコードブックにより生成され、該選択されたコードブックの表示は符号化情報に組み立てられることを特徴とする請求項1又は2に記載の方法。

【請求項4】 符号化情報からドキュメントを表す復号化情報を回復させる方法において、該ドキュメントは複数の要素からなり、

ドキュメントを表す符号化情報を受信し、その中の複数のコードを識別し、各々のコードは開始を有し、少なくとも各々の要素の部分を表し、各々の要素タイプを示す各々のタイプ表示を伝達し、各々の要素の1つ又はそれ以上のシンタックス的特徴を表すシンタックス情報の存在又は非存在を示す各々のシンタックス表示を伝達し、各々のコードの開始を基準とする所定の位置において各々のコードから各々のシンタックス表示を取得し、複数の復号化表示を生成し、各々の復号化表示は各々のコードから得られ、各々のコードにより表される各々の要素の部分に対応し、該各々のシンタックス表示はシンタックス情報を表す復号化表示の生成を制御し、該各々の復号化表示は、各々の復号化表示の情報の必要容量が各々のコードの情報の必要容量よりも大きい形で得られ、

ドキュメントを表す出力情報を生成するために、複数の復号化表示とコードにより表されない複数の要素の部分とを組み立てることを特徴とする方法。

【請求項5】 前記要素はタグベースのマークアップ言

語に準拠し、各要素はマークアップ言語タグからなり、シンタックス情報はタグ属性及びタグ内容を含むことを特徴とする請求項1ないし4のうちいずれか1項に記載の方法。

【請求項6】 前記タグベースのマークアップ言語は標準汎用マークアップ言語（SGML）ドキュメントタイプ定義（DTD）に準拠していることを特徴とする請求項5に記載の方法。

【請求項7】 前記コードは、シンタックス表示が要素タイプに独立にシンタックス情報の存在又は非存在を示す形を有することを特徴とする請求項1又は3ないし6のうちいずれか1項に記載の方法。

【請求項8】 請求項4又は請求項4と組み合わせられた請求項5ないし7のうちいずれか1項に記載の方法において、該方法は更に、

出力情報をその中の要素に従い処理することにより装置に表示するための表現を表す信号を生成し、該処理は、表現の1つ又はそれ以上の特徴に影響を与えるシンタックス情報の処理を回避するために、上記出力情報の中の1つ又はそれ以上の要素のシンタックス表示を使用することを特徴とする方法。

【請求項9】 請求項4又は請求項4と組み合わせられた請求項5ないし8のうちいずれか1項に記載の方法において、該方法は更に、

符号化情報をその中のコードに従い処理することにより装置に表示するための表現を表す信号を生成し、該処理は、表現の1つ又はそれ以上の特徴に影響を与えるシンタックス情報の処理を回避するために、上記符号化情報の中の1つ又はそれ以上のコードのシンタックス表示を使用することを特徴とする方法。

【請求項10】 請求項4又は請求項4と組み合わせられた請求項5ないし9のうちいずれか1項に記載の方法において、前記符号化情報は、サポートされていないコードであってそこからは各々の復号化表示が得られないコードの1つ又はそれ以上のインスタンスを含み、前記出力情報はまた該サポートされていないコードの1つ又はそれ以上のインスタンスを組み立てることにより生成されることを特徴とする方法。

【請求項11】 前記コードは固定長を有し、該コードの開始を基準とする固定位置にシンタックス表示を伝達することを特徴とする請求項1ないし10のうちいずれか1項に記載の方法。

【請求項12】 請求項4又は請求項4と組み合わせられた請求項5ないし11のうちいずれか1項に記載の方法において、前記符号化情報は、複数のコードブックから選択された選択コードブックの表示を含み、少なくともいくつかの復号化表示は該選択コードブックによるコードから得られることを特徴とする方法。

【請求項13】 ドキュメントを表す復号化情報を、圧縮された形式で複数の符号化要素からなる符号化情報か

ら回復させる方法において、該方法は、要素タイプを識別し、要素のシンタクスの特徴のシンタクス表示を得るために符号化要素を処理し、該シンタクス表示は、符号化要素の開始を基準とする符号化要素の中における所定の位置から得られ、要素タイプの圧縮表示はマークアップ言語タグの圧縮されない形に展開され、

シンタクス表示が、少なくとも1つのタグ属性が存在することを示す場合に、タグ属性情報の圧縮された表現をマークアップ言語のタグ属性名又はタグ属性値の非圧縮形に展開することにより、符号化要素の中のタグ属性情報を処理し、

シンタクス表示が、タグ内容が存在することを示す場合に、該タグ内容に適したプロセスにより、符号化要素の中のタグ内容情報を処理することを特徴とする方法。

【請求項14】 マークアップ言語タグは標準汎用マークアップ言語（SGML）ドキュメントタイプ定義（DTD）に準拠することを特徴とする請求項13に記載の方法。

【請求項15】 前記符号化要素は、シンタクス表示が要素タイプとは独立にシンタクス情報の存在又は非存在を示す形を有することを特徴とする請求項13又は14に記載の方法。

【請求項16】 前記出力情報をその中の要素に従い処理することにより装置に表示するための表現を表す信号を生成し、該処理は、表現の1つ又はそれ以上の特徴に影響を与えるタグ属性情報又はタグ内容の処理を回避するために、上記出力情報の中の1つ又はそれ以上の要素のシンタクス表示を使用することを特徴とする請求項13ないし15のうちのいずれか1項に記載の方法。

【請求項17】 前記符号化情報は、マークアップ言語タグの非圧縮形に展開されず、サポートされていないタイプの符号化要素の1つ又はそれ以上のインスタンスを含み、前記出力情報はまた該サポートされていない符号化情報の1つ又はそれ以上のインスタンスを組み立てることにより生成されることを特徴とする請求項13ないし16のうちのいずれか1項に記載の方法。

【請求項18】 前記符号化情報は複数のコードブックから選択された1つの選択コードブックの表示を含み、該選択コードブックに従い、前記マークアップ言語タグ、タグ属性名又はタグ属性値は非圧縮形に展開されることを特徴とする請求項13ないし17のうちのいずれか1項に記載の方法。

【請求項19】 要素タイプの前記圧縮表示は固定長を有し、前記シンタクス表示は要素タイプの圧縮表示の中の固定位置に伝達されることを特徴とする請求項13ないし18のうちのいずれか1項に記載の方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、一般的に非常に低

帯域の通信チャネル上で装置への伝送を行なうための情報の圧縮に関する。特に、本発明は、ハンドヘルドモバイル電話のようなワイヤレス装置への伝送のための、汎用マークアップ言語に準拠するドキュメント記述のようなシンタクス構造を有する情報の圧縮に関する。

【0002】

【従来の技術】 インターネットのようなネットワークは何年も存在している。しかしながら、それらは最近まで情報交換の一般的な媒体では無かった。インターネット使用の最近の爆発的成長は、例えば、大部分、サーバのネットワークを介して記録されたマルチメディア情報にアクセスしそれを精読するためにユーザがとらなければならないアクションを単純化する装置と方法の発展に依る。ハイパーリンクとして知られたリソースを参照することにより、異なる個々の情報が連続しない方法で組織化され、ユーザがそのリンク付けされた情報の中で容易にナビゲートできる。ユニフォームリソースロケータ

（URL）として知られる一意の識別子を、ネットワークを通して利用可能なマルチメディア情報の各々に割り当てることにより、どこに格納されているかにかかわらず、情報にすぐにアクセスすることができる。"ハイパーメディア" ネットワークにおけるネットワーククライアント及びサーバはここではそれぞれハイパーメディアクライアント及びハイパーメディアサーバと称される。

【0003】 この成長に貢献した1つの重要な発展は、"マークアップ言語" などのような機構の使用、及び、ドキュメントの種々のシンタクスの特徴を特定する広範囲のエLEMENTを定義し実装する関連のプロセスである。今日使用されている多くのマークアップ言語は国際標準ISO8879:1986に準拠し、その標準は、ここで標準汎用マークアップ言語（SGML（Standard Generalized Markup Language））と呼ばれるタグベースの言語のための基本的な規則のセットを定義する。おそらく、SGMLに準拠してインターネット上で最も広く使用されるマークアップ言語はハイパーテキストマークアップ言語（HTML）である。

【0004】 タグベースのマークアップ言語により表されるドキュメントは、典型的にはブラウザ又はリーダーと称されるソフトウェアアプリケーションにより表示され操作される。これらのソフトウェアアプリケーションは、ドキュメントが正しく表示されるようにドキュメントを表す情報を解剖し解釈するための、適切なマークアップ言語規則に準拠するプロセスを実装している。

【0005】 SGMLのようなマークアップ言語によるドキュメントを表す情報は、タグ及び関連するタグ属性及びタグ内容を有するいくつかの要素からなる。これらの要素は、ドキュメント中の伝達される情報のシンタクスの特徴を伝達する。タグは要素タイプを識別する。例えば、HTMLにおいて、ドキュメント全体を表す要素は、ドキュメントの開始と終わりをマークするタグに

より識別される。テキストのパラグラフを表す要素はパラグラフの開始をマークするタグにより識別され、アンダーラインが付いて表示されるテキストはアンダーラインの開始及び終わりをマークするタグにより識別される。

【0006】タグ属性は要素の1つ又はそれ以上の特徴を特定する情報を提供する。ドキュメントに埋め込まれるイメージファイルを表すタグは、例えば、埋め込まれるイメージファイルの名前を特定する属性を含む。マークアップ言語の仕様により、タグ属性はオプションでもよく、又は、関連するタグタイプにより要求されてもよい。

【0007】タグ内容は、一般的に表示されるか又はそうでなければユーザにより操作可能な情報を表す。タグ内容はオプションであり得、又は、タグのタイプにより要求され得、そして、タグ、属性及び内容を有する他の”入れ子”になった要素を含み得る。SGMLに準拠するようなマークアップ言語は、ドキュメントの要素をインプリメントするための非常に柔軟でパワフルな機能を提供できる。なぜならばSGML自身が非常に柔軟だからである。この柔軟性はコストなしでは得られない。追加の帯域がタグとタグ属性とを運ぶために必要となり、追加のリソースがタグ及びタグ属性を解剖し解釈するために必要となる。HTMLにおいて、タグ及び属性は<tagid name=value>と同様の形態における文字列により表現される。ここで、tagid はタグ識別子であり、nameは属性の名前であり、value はその属性に付与される値である。タグは1つ以上の属性を持ち得る。

【0008】

【発明が解決しようとする課題】タグ及びタグ属性を伝達し、処理するために必要な追加の帯域及びリソースは多くの場合に重大な欠点とはならない。なぜならば、十分なコンピューティングパワーを有するパーソナルコンピュータ及び他のワークステーション及び十分な帯域を有する通信チャンネルは既に使用可能だからである。

【0009】しかしながら、特にワイヤレス電話機のようなハンドヘルド装置等のモバイル装置を介する、インターネットのようなネットワークと接続したハイパーメディアサーバへのアクセスを提供することへの興味は増加している。これらの装置は処理能力及びメモリスぺースにおける厳しい制限により特徴付けられる。更に、モバイル装置をネットワークの残りに接続する通信チャンネルの帯域はまた厳しく制限されている。

【0010】ワイヤレス電話機はまた典型的なデスクトップ又はポータブルコンピュータにより供給されるリソースの少しの部分しか持たない。典型的には、その処理能力は多くのコンピュータの処理能力の1パーセント以下であり、メモリスぺースは一般的に150キロバイト(kB)よりもずっと小さい。通信パスは普通400から19200ビット毎秒の範囲にあり、その通信パスを

用いるコストは100kBもしくはそれ以上あたりの米国ドルに換算して計測される。

【0011】これらの通信チャンネルの制限された帯域はこれらのチャンネルで運ばれる情報の要求容量を減少させることにより、より効率的に使用され得る。情報容量の要求量はデータもしくは情報の圧縮のいくつかの形態を採用することにより減少させ得る。ホフマン符号化のような一般目的の圧縮スキームが考えられてきたが、残念なことに、一般目的のスキームは魅力的でない。なぜならば、結果として得られる圧縮情報は、背後にある情報のシンタックスの特徴を包み隠すからである。言い換えれば、タグを識別すること及びタグの属性とタグの内容の存在または非存在が、圧縮表現からは簡単には決定され得ない。更に、一般目的の圧縮スキームは、通常、特定のマークアップ言語に基く圧縮スキームほどには情報容量要求量を減らすことができない。

【0012】HTMLのような特定のマークアップ言語に基く種々の圧縮スキームが考えられてきた。この種の圧縮スキームは、特定のマークアップ言語の知られた特徴を利用することにより、高レベルの圧縮を達成することを可能とする。例えば、マークアップ言語特有の圧縮スキームは、内容を持ち得ないタグのタグ内容を運ぶ可能性を許容する必要はない。不運にも、これらのスキームは、ブラウザ又は展開プロセスが全ての圧縮要素を処理又は展開することを要求する。マークアップ言語への展開又は変更は、ブラウザが新しい言語の機能処理するために修正されない限り、圧縮表現から回復することはできない。さもなければ、新たな機能の圧縮は、入れ子になった要素とともに新たな機能を取り入れたそれらの要素のシンタックスの特徴を不明瞭にする。新たな機能を使用できないか又は使用する必要のないアプリケーションや装置に取り入れられているとしても、ブラウザは相当に変更されなければならない。

【0013】例えば、マークアップ言語ベースの圧縮スキームが新たな表示フォーマットを圧縮するように拡張される場合、ブラウザが新たな機能を展開するために必要とされる処理を含むように修正されなければ、圧縮表現からその表示フォーマット情報を回復できない。更に、そのような修正がなければ、ブラウザはその新たな機能を見逃したりスキップすることができず、残りの情報を展開することができない。なぜならば、その処理能力は圧縮された新たな機能の範囲を決定できないからである。

【0014】本発明の目的は、背後にあるドキュメント要素のシンタックスの特徴を不明瞭にせず、ドキュメントを表現する情報を伝達して処理するために必要な帯域及びリソースを減らすことである。

【0015】

【課題を解決するための手段】本発明の一側面は、ドキュメントを表す入力情報の容量の必要量を減少させる方

法であって、入力情報を受信し、複数の要素を識別し、各要素は各々のタイプを有し、少なくともいくつかの要素は1つ又はそれ以上の各々のシンタックス的特徴を表すシンタックス情報を有し、複数のコードを生成し、各々のコードは開始を有し、表された部分の情報の必要容量よりも低い必要容量を有する形で少なくとも各々の要素の部分を表し、該各々のコードは、各々の要素タイプと、該各々の要素のシンタックス情報の存在又は非存在を示すシンタックス表示とを伝達し、該各々のコードは該各々のコードの開始を基準とした所定の位置に該シンタックス表示を伝達し、複数のコードと該複数のコードでは表されない複数の要素の部分とを伝送又は格納に適する形に組み立てることによりドキュメントを表す符号化情報を生成する方法である。

【0016】本発明の他の側面は、符号化情報から複数の要素からなる情報を回復させる方法であって、ドキュメントを表す符号化情報を受信し、その中の複数のコードを識別し、各々のコードは開始を有し、少なくとも各々の要素の部分を表し、各々の要素タイプを示す各々のタイプ表示を伝達し、各々の要素の1つ又はそれ以上のシンタックス的特徴を表すシンタックス情報の存在又は非存在を示す各々のシンタックス表示を伝達し、各々のコードの開始を基準とする所定の位置において各々のコードから各々のシンタックス表示を取得し、複数の復号化表示を生成し、各々の復号化表示は各々のコードから得られ、各々のコードにより表される各々の要素の部分に対応し、該各々のシンタックス表示はシンタックス情報を表す復号化表示の生成を制御し、各々の復号化表示の情報の必要容量が各々のコードの情報の必要容量よりも大きい形で得られ、ドキュメントを表す出力情報を生成するために、複数の復号化表示とコードにより表されない複数の要素の部分とを組み立てる方法である。

【0017】本発明の更なる他の側面は、ドキュメントを、圧縮された形式の複数の符号化要素から回復させる方法であって、要素タイプを識別し、要素のシンタックス的特徴のシンタックス表示を得るために符号化要素を処理し、該シンタックス表示は、符号化要素の開始を基準とする符号化要素の中における所定の位置から得られ、要素タイプの圧縮表示はマークアップ言語タグの圧縮されない形に展開され、シンタックス表示が、少なくとも1つのタグ属性が存在することを示す場合に、タグ属性情報の圧縮された表現をマークアップ言語のタグ属性名又はタグ属性値の非圧縮形に展開することにより、符号化要素の中のタグ属性情報を処理し、シンタックス表示が、タグ内容が存在することを示す場合に、該タグ内容に適したプロセスにより、符号化要素の中のタグ内容情報を処理する方法である。

【0018】本発明の種々の特徴及びその好ましい実施形態は以下の説明及び添付の図面を参照することにより、より良く理解され得、いくつかの図面において参照

数字は同一の要素を参照する。以下の説明及び図面の内容は例としてのみ説明され、本発明の範囲における限定を示すものと理解するべきでない。

【0019】

【発明の実施の形態】(概要) 図1は本発明の種々の側面が実行され得るシステムの主要な構成要素を示す概要図である。同図に示されるいくつかの構成要素は種々の実施形態において省略され得る。同図に示すように、クライアント1はサーバ51及びサーバ52により提供されるリソースにアクセスするためにネットワーク40を使用する。サーバ51及びサーバ52はハイパーメディアサーバであり、おそらくハイパーテキストトランスファプロトコル(HTTP)に従って動作すると考えられるが、これは本発明の実施に必要なことではない。典型的な実施形態において、リモート装置11は、ユーザに情報が表示され得、ユーザからの入力が受け入れられるユーザインターフェースを提供する。そして、コンピュータ31は、従来のネットワーククライアントと同一の方法でネットワーク40と情報を交換する。

【0020】コンピュータ31は記憶装置32にパラメータ及び情報を格納する。記憶装置32は典型的には、ランダムアクセスメモリ(RAM)、リードオンリーメモリ(ROM)及び、磁気及び光ディスクドライブのような長期記憶装置の組み合わせである。コンピュータ31は受信機21及び送信機22を介してリモート装置11と通信を行なう。送信機22を介してコンピュータ31により送信された情報は受信機16を介してリモート装置11により受信される。送信機15を介してリモート装置11により送信された情報は受信機21を介してコンピュータ31により受信される。

【0021】図1に示す実施形態において、リモート装置11は、ディスプレイ12、1又はそれ以上のボタン13、記憶装置14、送信機15及び受信機16からなる。例えば、装置11は、Mitsubishi Wireless Communications, Inc. のMobileAccessTM電話機、もしくはSamsung Electronics Corporation のDuette電話機のようなワイヤレス電話機であり得る。典型的なワイヤレス電話機において、ディスプレイ12は液晶ディスプレイ(LCD)パネルである。ボタン13は、スイッチ、キー、ボタンのような1又はそれ以上のデータ入力装置を表す。記憶装置14はデジタル情報を格納できるメモリ回路又は他の装置を表す。好ましくは、少なくとも記憶装置14の1部は永続的な記憶装置であり、装置11がオフとなったときでもそこに情報を保持する。いくつかの実施形態においては、記憶装置14の1部は統合されたプッシュ/プルキャッシュに構成される。また、記憶装置14の1部は、永続的メモリ又はROMにプログラム命令を格納し、装置11はマイクロプロセッサ又はプログラム命令を実行できる他の種類の処理回路からなるということが考えられる。

【0022】コンピュータ31、サーバ51、サーバ52、受信機21及び送信機22の間の通信パスの種類は本発明の実施には重要でなく、例えば、プライベート及び／又はパブリックな設備を使用する交換及び／又は非交換パスとして実装され得る。同様に、ネットワーク40のトポロジは重要でなく、階層型及びピアツーピア型のネットワークを含む種々の方法で実装され得る。コンピュータ31及びサーバ51は互いにローカルな位置に設置され得、また、同一ハードウェア上にも実装され得る。

【0023】コンピュータ31と装置11間の通信パスの種類もまた本発明の実施においては重要でない。しかしながら、多くのアプリケーションでは、装置11は、無線周波数からスペクトルの赤外線部分での電磁的伝送のような通信技術を使用するワイヤレス装置である。装置11が例えばセルラー電話機のようなワイヤレス電話機である場合のアプリケーションにおいて、送信機15、受信機16、受信機21及び送信機22は通常の電話呼のために使用される通信設備を表す。

【0024】(リモート装置) リモート装置11及びコンピュータ31がHTTPクライアントとしてクライアント1を実装する場合のアプリケーションにおいて、装置11は少なくとも3つの基本的な機能を提供する。

(1) ナビゲーション機能がユーザにHTTPのユニフォームリソースロケータ(URL)ハイパーリンクをナビゲート又は渡らせ、(2) 通信機能がコンピュータ31と情報を交換し、(3) インターフェース機能は、情報がユーザに表示され得、ユーザから入力を受け入れられ得るユーザインターフェースを提供する。

【0025】好ましくは、これらの機能は事象駆動アーキテクチャを使用するソフトウェア制御プロセスにより実装される。事象は例えばボタン13を介してユーザにより開始され得、もしくは、受信機16を介して受信した信号により開始され得る。ナビゲーション機能は2つの状態のどちらかで動作する。”レディ”状態において、装置は行こうとしているハイパーリンクを特定するユーザ入力を待つ。”ペンディング”の状態において、通信機能はコンピュータ31に要求を送信し、装置はコンピュータ31からの応答を待っている。HTTPに関して、レディ状態ではハイパーメディア要素のURLを表示又は処理するために特定するユーザ入力を待ち、ペンディング状態ではコンピュータ31が要求されたハイパーメディア要素を提供することを待つ。

【0026】ある実施形態によれば、ハイパーメディア情報はハンドヘルドデバイストランスファプロトコル(HDTP)によりコンピュータ31と交換される。このプロトコルの1つのバージョンは、Unwired Planet, Inc., Redwood Shores, Californiaにより1997年7月15日に発行された”HDTP仕様”のパート番号HDTP-SPEC-DOC-101に説明されており、参

照として全体的にここに取り入れられている。HDTPはHTTPに似ているがワイヤレス電話のようなリモート装置での使用に最適化され、好ましくはユーザデータプロトコル/IP(UDP/IP)を使用して運ばれる。UDP/IPは、例えば、パケットが受信されることを保証せず、パケットが送られた順番で受信がなされることを保証しないという理由で、一般的にTCP/IPよりも信頼性が低いと考えられている。しかしながら、UDP/IPのようなデータグラムプロトコルは本発明を実施する上では魅力的である。なぜならば、UDP/IPは、情報が交換される前に、送信者及び受信者間で”コネクション”を確立することを要求しないからである。このことは、セッション生成時に大量のパケットを交換する必要を削減させる。

【0027】好ましい実施形態において、ハイパーメディア情報は、ハンドヘルドデバイスマークアップ言語(HDML)によりカード及びデッキに構成される。多重デッキ及び他の種類のメッセージ要素はダイジェストと呼ばれる情報構造に構成される。このマークアップ言語のあるバージョンはUnwired Planet, Inc.により1997年3月に発行された”HDML 2.0仕様”のパート番号HDML-SPEC-DOC-200、改定A、に説明されている。それは全体的にここに参照として取り入れられている。

【0028】(中間のコンピュータ) ここで議論された実施形態によれば、コンピュータ31は、リモート装置11とともに従来のハイパーメディアクライアントの機能を提供する。この実施形態において、コンピュータ31はHDTPによりリモート装置11から情報を受信し、必要に応じてそのHDTP情報を対応するHTTP情報に翻訳し、その結果をサーバ51に送信する。同様に、コンピュータ31はHTTPによりサーバ51から情報を受信し、必要に応じてそのHTTP情報を対応するHDTP情報に翻訳し、その結果をリモート装置11に送信する。コンピュータ31とリモート装置11の間で交換されるHDTP情報は、情報の要求容量を削減させ、情報を解剖し、解釈するためにリモート装置11に必要とされる処理を削減させるために本発明により圧縮される。この圧縮及び相補的な展開は、リモート装置11及びコンピュータ31において実行される符号化及び復号化のプロセスにより実行される。

【0029】(プロセス) 図2は、ドキュメント要素の圧縮された表現を生成するための本発明による符号化プロセスの一実施形態を示す図である。識別要素62はドキュメントを表す情報をパス61から受信し、その情報内で複数の要素を識別する。要素は典型的には、少なくとも構造のいくつかの側面及びドキュメントのシンタックス的な特徴を表すシンタックス情報を有する。

【0030】符号化64は、少なくともいくつかのドキュメント要素の少なくとも一部分を表す複数のコードを

生成する。少なくともいくつかのコードは、表現される要素情報よりも低い情報容量が必要である。そのコードは、要素のためのシンタックス情報が存在するかどうかの表示とともに表された要素のタイプを伝達する。好ましくは、少なくともいくつかのシンタックス情報が、情報の要求容量を低下させる方法で符号化される。入れ子になった情報を処理する必要に応じて、要素情報はパス63で渡される。入れ子になった情報は再帰的処理を含む種々の方法で処理され得る。

【0031】組み立て66は、符号化64により生成されたコード、及び、それらのコードにより表されない要素又は要素の部分、送信もしくは格納に適した形に組み立てることにより、ドキュメントを表す符号化された情報をパス67に沿って生成する。本発明の他の実施形態は、複数のコードブックを提供するコードブック68を含む。符号化64は、適合するように、この複数のコードブックから1つのコードブックを選択し、選択されたコードブックにより1又はそれ以上のコードを生成する。選択されたコードブックの表示は、符号化された情報に含まれる。

【0032】図3は、ドキュメントの要素を符号化表現から回復させるための本発明による復号化プロセスの一実施形態を示す。識別要素72はドキュメントを表す符号化された情報をパス71から受信し、各々が少なくともそれぞれのドキュメント要素の1部分を表す複数のコードを識別する。そのコードに回答して、復号化74はシンタックス表示を取得し、復号化された表現を生成する。少なくともいくつかの復号化表現は、対応するコードよりも大きな情報容量を必要とする。シンタックス表示は、ドキュメントの1つ又はそれ以上のシンタックスの特徴を表すシンタックス情報が存在するしないを示す。復号化された表現は必要に応じてパス73に沿って渡され、入れ子になったコードを扱うために処理される。入れ子になったコードは、再帰的プロセスを含む種々の方法で処理され得る。

【0033】組み立て76は、復号化74により生成された復号化表現、及び、それらのコードにより表されない要素又は要素の部分を組み立てることにより、ドキュメントを表す出力情報をパス77に沿って生成する。本発明の他の実施形態は、複数のコードブックを提供するコードブック78を含む。復号化74は、符号化情報の中の選択されたコードブックの表示に回答してこの複数のコードブックから1つのコードブックを適切に選択し、選択されたコードブックに従って1つ又はそれ以上の復号化表現を生成する。

【0034】本発明の他の実施形態において、処理80はパス77から出力情報を受信し、表示のためのプレゼンテーションを表す信号をパス81に沿って生成する。ある状況では、復号化74は、コードが知られておらず又は復号化プロセスにおいてサポートされていないとい

う理由で復号化できないコードと出会うかもしれない。復号化74はこれらのサポートされていないコードを、そのコードを使用することができる次のプロセスのためにパス73で渡し得る。処理80はそれらのコードの処理をスキップもしくは回避するためにサポートされていないコードのシンタックス表示を使用する。

【0035】更なる実施形態において、復号化74は、ディスプレイのための表現を表す信号を生成するためのプロセス80と同様のプロセスを含む。この実施形態において、例えば、ディスプレイ装置がコードにより表された要素に適切に回答できないために、どのコードがスキップされるべきかを決定するために、復号化74におけるこのプロセスは、コードの中に伝達される要素タイプ及びシンタックス表示を使用する。

【0036】(符号化)

(状態プロセス) 符号化64の符号化プロセスは図4に示されるような状態プロセスで説明され得る。状態の各々は円で表されている。状態間の遷移は線で表され、矢印で示された方向で遷移が起こる。

【0037】符号化プロセスは状態100(開始)で始まり、110に沿って状態110(符号化タグ)に遷移する。状態101は各要素タグの符号化表現を生成する。各要素タグが関連するシンタックス情報を伴っていない場合、パス111に沿って遷移し、続く要素タグの符号化表現を生成する状態101に遷移する。1つもしくはそれ以上のタグ属性が存在する場合、パス112に沿って状態102(符号化属性名)に遷移する。タグ属性が存在せずにタグ内容がある場合、パス118に沿って状態105(符号化内容)に遷移する。更なる要素タグが存在しない場合、符号化プロセスを終了するために、パス122に沿って状態107(終わり)に遷移する。

【0038】状態102は、各属性名の符号化表現を生成する。パス113に沿って状態103(符号化属性値)に遷移する。状態103は対応する属性値の符号化表現を生成する。次のタグ属性が存在すれば、114に沿って、次のタグ遷移のための符号化表現を生成する状態102に遷移する。更なるタグ属性が存在しないとき、パス115に沿って状態104(属性終わり)に遷移する。状態104は、各要素のためのタグ属性の終わりを知らせるコードを生成する。タグ内容が存在する場合、タグ内容を処理するために117に沿って状態105に遷移する。内容が存在しない場合、次の要素タグを処理するためにパス116に沿って状態101に遷移する。

【0039】状態105は各タグ内容の符号化表現を生成する。次のタグ内容が存在する場合、次の内容を処理するためにパス119に沿って状態105に遷移する。更なるタグ内容が存在しないとき、パス120に沿って状態106(内容終わり)に遷移する。状態106は各要素のタグ内容の終わりを知らせるコードを生成する。

次の要素タグを処理するためにパス121に沿って状態101に遷移する。

【0040】以下でより詳細に説明するように、タグ内容は入れ子の要素を含み得る。入れ子の要素が存在する場合、状態100への再帰的遷移が図示はされていないパスに沿って行なわれる。入れ子の特定のレベルにおける全ての要素が処理されると、状態105への再帰的戻り遷移が図示はされていない他のパスに沿ってなされる。

【0041】(例) 図7はHTMLのようなマークアップ言語で表現されるシンプルなドキュメントを示す。ドキュメントは行で並べられており、各行はこの説明で参照が便利のように番号付けされている。行番号はマークアップ言語の部分で形成していない。実際の実施形態においては、マークアップ言語により提供されるもの以外には、ドキュメントは行の表示や他の区別無しに伝達され得ることが考えられる。

【0042】行1は、HTMLドキュメントの始まりをマークする<HTML>タグを含み、行8はドキュメントの終わりをマークする</HTML>タグを含む。この例において、<HTML>タグは属性を持たないが、内容を持ち、それは行2及び行7における始まり及び終わりのBODYタグによりマークされるドキュメントの本体である。<BODY>タグは属性を持たないが、内容を持つ。<BODY>タグの内容は<HTML>タグ内で入れ子になっている。BODYタグの内容は、行3から6に示されており、テキストといくつかのタグからなる。

【0043】行3に示される<BODY>タグの内容の部分はシンプルなテキストを表している。行4に示される内容の部分は、IMGタグを有する要素であり、内容は持たないが、表示のためのイメージのソースを特定する名前(src)と値("/item.gif")を有する属性を有する。行5に示される内容の部分は、ボールド体のフォントで表示する単語をマークする開始と終わりのBタグを有する要素のペアを含む。どちらのタグも属性は持たないが、テキスト内容を有する。行6に示される内容の部分は、開始及び終わりのAタグを有する要素を含むテキストである。<A>タグは両方とも属性及び内容を持っている。タグ属性は、他のドキュメントのURLを特定する名前(href)及び値("http://a.url/info")を有する。<A>タグの内容は終わりのタグの直前に現れるテキスト"here"である。

【0044】図8は、上述した符号化プロセスを、図7に示したドキュメントマークアップ言語に適用することにより得られた符号化表現を示す図である。図8に示した符号化表現は、説明において参照に便利のように番号付けされて並べられており、分かりやすくするためにインデントが付けられている。実際の実施形態では、マ

ークアップ言語要素の符号化表現により提供されたものの以外は、行の表示や他の区別は含まない形で符号化情報が生成されることが考えられる。

【0045】図8を参照すると、表記{XYZ-AC}はマークアップ言語タグ<XYZ>の符号化表現を含み、1つもしくはそれ以上の属性が存在することの表示とタグ内容が存在することの表示を含むコードを表す。例えば、行1において、表記{HTML-C}は、<HTML>タグの符号化表現を含み、タグ属性はないがタグ内容は存在することの表示を含むコードを表す。同様に、行4.1における表記{IMG-A}は、タグの符号化表現を含み、1つ又はそれ以上のタグ属性は存在するが、タグ内容は存在しないことの表示を含むコードを表す。

【0046】図8に示す例によれば、行1に示す表記{HTML-C}は、図7の行1に示す<HTML>タグを表すコードを表す。上述したように、コードは、表された要素タイプを伝達し、タグ内容が存在することの表示を含む。行2において、表記{BODY-C}は、<BODY>タグ(行2、図7)を表し、タグ内容が存在することを示すコードを表す。

【0047】行3において、表記{STR}は、テキストの存在をマークするために使用される特別なコードを表す。表記"The item"はテキストそのものを表している。このコードはいつもタグ内容が存在することを暗黙のうちに示している。テキストは明白もしくは暗黙的なコードを使用する種々の方法でマークされ得る。例えば、テキストストリングの始まりは、テキスト文字のためにある値をリザーブしておくことにより暗黙的にマークされ得る。これらのリザーブされた値は例えばバイナリデータのフィールドで発生しやすいので、そのようなスキームは一般的に文脈依存である。好ましい実施形態において、テキストストリングの始まりは、図で示された表記{STR}により表されるような明白なコードでマークされる。テキストストリングの終わりは、ヌルやバイナリ0のような特別の文字により明白に、始まりのコードとともに含まれる表示長さ値により明白に、又は、有効なテキスト文字ではないコードにより暗黙的にマークされ得る。本発明の実施において特定のスキームが重要であることはない。

【0048】行4.1から4.3は、図7の行4に示すドキュメント要素の符号化表示を一括して表している。行4.1において、表記{IMG-A}はタグを表し、1つ又はそれ以上のタグ属性が存在することを示すコードを表す。行4.2において、表記{src}はタグ属性の名前"src"を表すコードを表す。このコードは、以下で完全に説明する名前自身の圧縮表現、もしくは、属性名が従来からあるテキストストリングのような他の形で特定されることを示す包括的な属性コードであり得る。表記("/item.gif")

は、属性の値を提供する従来からあるテキストストリングを表す。一方、属性値はバイナリコードのようないくつかの他の形式に符号化され得る。行4.3において、表記{END:img-a}はタグのためのタグ属性の終わりをマークするコードを表す。本発明の一実施形態において、1つのコードが属性の終わりをマークするために使用され、他のコードが内容の終わりをマークするために使用される。他の実施形態において、異なるコードは要素タイプに従って使用される。さらに他の実施形態において、異なるコードは属性及び要素タイプによる内容の終わりをマークするために使用される。行4.3に示す例を参照すると、これらの実施形態により、表記"img-a"は、IMG属性の終わりをマークするための一意の{END}コードを表すものと理解され得る。しかしながら、好ましい実施形態において、ヌルや0の値のような1つの特定のコードが全てのタイプのタグの属性と内容の終わりをマークするために使用される。この実施形態では、表記"img-a"は、単に、タグ属性のコード、内容、終わりのコード間の一致を読者に示す便宜にすぎないということが理解され得る。

【0049】行5.1から5.5は、図7の行5に示したドキュメント内容の符号化表現を一括して表している。行5.1及び行5.3において、表記{STR}及び付随するテキストは、コード及び図7の行5に示す2つのテキストストリングを表すテキストを意味する。行5.2.1から5.2.3は、図7の行5における第1の要素の符号化表現を一括して表す。行5.2.1において、表記{B-C}はタグを表し、内容が存在することを示すコードを意味する。行5.2.2において、テキスト内容は上述した表記{STR}"red"により表される。行5.2.3において、表記{END:b-c}は、タグの内容の終わりをマークするコードを意味する。同様に、行5.4.1から行5.4.3は、図7の行5における第2の要素の符号化表現を一括して示す。

【0050】行5.5において、表記{STR}"for a limited time (制限された時間のために)"は上述したテキストストリングの符号化を表し、図7の行5に示すドキュメント内容の符号化表現が完了する。図8に示した例によれば、図7の行6におけるドキュメント内容の符号化表現は行6.1.1から6.3に一括して示される。しかしながら、本発明の現実の実施形態では、隣接するテキストストリング"for a limited time"及び"Click"は、{STR}"for a limited time. Click"により表される1つの符号化表現に結合され得る。

【0051】ちょうど説明したように、行6.1.1から6.3は図7の行6に示すドキュメント内容の符号化

表現を一括して表す。説明したように、行6.1.1の表記{STR}"Click"はテキストストリングの符号化を表す。行6.1.2において、表記{A-A-C}は、<A>タグを表し、タグの属性と内容が存在することを示すコードを意味する。行6.1.3において、表記{href}("http://a.url/info")は、タグ属性の名前と値を表す符号化を表す。行6.1.4において、表記{END:a-a}は、<A>タグの属性の終わりをマークするコードを表す。行6.2.1において、表記{STR}"here"は、タグの内容であるテキストストリングの符号化を表す。行6.2.2において、表記{END:a-c}は、タグ<A>の内容の終わりをマークする。行6.3における表記はテキストストリングを表し、図7の行6に示すドキュメント内容の符号化表現を完了する。

【0052】行7において、表記{END:body-c}及び{END:html-c}は、それぞれ<BODY>及び<HTML>タグの内容の終わりをマークするコードを表す。

(圧縮)種々の符号化もしくは圧縮スキームが、コードにより表されるドキュメント要素、もしくはドキュメント要素の部分の情報容量の必要量よりも低い情報容量の必要量を有するコードを生成するために使用され得る。そのコードは、表されているドキュメント要素のタイプ、及びドキュメント要素にシンタックス情報が存在するかどうかの表示の両方を伝達するために生成される。シンタックス情報の表示は、コードの始まりを基準とする所定の位置に伝達される。

【0053】本発明の好ましい実施形態によれば、コードは、1又はそれ以上のビット、すなわち2つの最上位ビットが、シンタックス情報が存在するかしないかを示すためにリザーブされる、固定長1バイト(8バイナリビット)を有する。例えば、HTMLにおいて、2つのビットは、1つ又はそれ以上のタグ属性及びタグ内容が存在するしないを示すためにそれぞれリザーブされ得る。他のコード構造は、可変長のコードを含むことが可能である。例えば、ホフマン符号化により生成された要素タイプの可変長表示、及びシンタックス情報の別の表示を含み得る。シンタックス情報の表示は、コードの開始を基準とするいかなる所定の位置にも置かれ得る。

【0054】要素タイプに応じて所定の位置を変えることを許容するルールが確立され得る。例えば、その位置は、要素タイプの可変長表示のすぐ後に続くように定義し得る。他の例として、1つの位置は、いくつかの特定の値の1つを有する特別なコードのクラスについて予め定め得、他の位置は他のコードについて定め得る。所定の位置は、好ましい実施形態においては、要素タイプと独立して固定である。

【0055】(特別なコード)好ましい実施形態におい

て、6つの特別のコードのクラスが設定される。これらの特別のコードは"global codes (グローバルコード)"と参照される。なぜならば、この実施形態によれば、全ての符号器及び復号器はこれらのコードを正しく解釈し処理できなければならないからである。これらの6つのコードは以下で説明される。

【0056】{CBK}と表記される特別のコードは、複数のコードブックから適切に選択されたコードブックを特定する値をマークする。復号化は、その選択されたコードブックに従い実行される。上記で簡単に説明したように、固定長8ビットコードが、要素タイプ及びシンタックス情報の表示の両方を伝達するために使用される。2つのビットがシンタックス情報の表示を伝達するために使用される場合、6ビットだけが要素タイプを伝達するために残る。一般的に、要素の数は、6ビットで表現可能なものをはるかに超える。これらのコードを頻繁に使用される属性名及び／又は属性値にも使用することが望ましいので、この制限は一層厳しくなる。コードを複数のコードブックに組織化し、この複数から適切なコードブックを選択することにより、コーディングスペースの大きさは相当に拡張され得る。符号器がコードブックを選択すると、選択の表示は符号化情報に組み立てられ、その結果、相補的な復号器は、どのコードブックが復号化に使用されるべきかを決定できる。{CBK}コードがこの表示をマークするものである。

【0057】{CHR}と表記される特別のコードは、文字を特定する値をマークする。例えば、情報交換のためのアメリカ標準コード(ASCII)に準拠したテキストにより表されるドキュメントは、ユニコードテキストにおいて定義されたいくつかの文字を表せない。どのようなユニコード文字も{CHR}コードによりマークされた数値により表される。

【0058】{DAT}と表記された特別のコードは、復号器により処理されない"opaque (不明瞭な)"データの開始をマークする。データは、データの内部構造が符号器に知られる必要がないという意味で不明瞭であると言える。不明瞭なデータはマークされ、修正無しに符号化された情報に含まれる。不明瞭データの範囲は{DAT}コードに付随する長さ値により伝達される。

【0059】{END}と表記された特別のコードは、上述したように、ある要素及びシンタックス情報の終わりをマークする。{STR}と表記された特別のコードは、上述したように、テキストストリングの開始をマークする。{UNK}と表記された特別のコードは、知られていない要素のタイプをマークする。このコードの使用は、符号器及び復号器が実装されたときには未定義であった要素を含むドキュメントを処理するために、現存する符号器及び復号器の能力を改善する。古いほうの符号器は、より最近の符号器に新しい要素を受信させて処

理させる形で、未知の要素を渡すことができる。古い符号器とともに動作する古いほうの復号器は、{UNK}コードによりマークされた要素をスキップし、他の知られたコードの処理を再開することができる。

【0060】(復号化)符号化74の符号化プロセスは、図5に示したような状態遷移プロセスに関して説明され得る。各状態は円で表される。状態間の遷移は線により表され、矢印で示す方向に遷移が行なわれる。復号化プロセスは状態130(開始)で始まり、140に沿って状態131(復号化タグ)に遷移する。状態131は、各コードから得られる各要素タグの復号化表現を生成する。各コードが、シンタックス情報が存在しないことを示す場合、パス141に沿って、次のコードから得られる復号化表現を生成する状態131に遷移する。コードが、1又はそれ以上の属性が存在することを示す場合、パス142に沿って状態132(復号化属性名)に遷移する。コードが、タグ属性は存在しないがタグ内容が存在することを示す場合、パス148に沿って状態135(復号化内容)に遷移する。更なる要素タグが存在しない場合、復号化プロセスを終了するために、パス152に沿って状態137に遷移する。

【0061】状態132は各属性名の復号化表現を生成する。パス143に沿って状態133(復号化属性値)に遷移する。状態133は、対応する属性値の復号化表現を生成する。次のタグ属性が存在する場合、144に沿って、次のタグ属性の復号化表現を生成する状態132に遷移する。更なるタグ属性が存在しないとき、タグ内容があれば、タグ内容を処理するために147に沿って状態135に遷移する。内容が存在しなければ、パス146に沿って、次のコードを処理するために状態131に遷移する。

【0062】状態135は各タグ内容の復号化表現を生成する。次のタグ内容が存在すれば、パス149に沿って、次の内容を処理するために状態135に遷移する。更なるタグ内容が存在しないとき、パス151に沿って、次のコードを処理するために状態131に遷移する。上述したように、タグ内容は入れ子のコードを含み得る。入れ子のコードが存在すれば、状態130への再帰的遷移が図示されていないパスに沿って行なわれる。入れ子の特定のレベルにおける全てのコードが処理されたとき、状態135への再帰戻り遷移が図示されていない他のパスに沿って行われる。

【0063】(再帰)図4及び図5に示した状態遷移図は再帰のための処理は示していない。再帰は本発明の実施には要求されていないが、入れ子の要素及びコードを処理するために多くの実施形態において効率的な技術である。図6に示される機能的フロー図は、HTMLのようなマークアップ言語で表現されるドキュメント要素の符号化もしくは復号化のための再帰的処理を表す。

【0064】(符号化)図示された符号化のプロセスに

よれば、ステップ221では種々の初期化タスクを実行する。ステップ222では再帰レベルを0に初期化する。ステップ223では符号化表現を生成するために要素タグを処理する。ステップ224ではタグ要素が存在するかどうかを調べる。存在する場合、ステップ225において符号化表現を生成するためにタグ属性を処理し、他のタグ属性が存在するかどうかを調べるためにステップ224に戻る。更なるタグ属性がないとき、プロセスはステップ226から続く。

【0065】ステップ226ではタグ内容が存在するかどうかを調べる。存在する場合、ステップ227で符号化表現を生成するためにタグ内容を処理する。ステップ228では要素がタグ内容の中で入れ子になっているかどうかを調べる。なっていないければ、他のタグ内容が存在するかどうかを調べるためにプロセスはステップ226に戻る。更なるタグ内容が存在しないとき、プロセスは230から続く。要素がタグ内容の中で入れ子になっている場合、ステップ229では再帰レベルをインクリメントし、プロセスはステップ223から続く。

【0066】ステップ230では現在の再帰レベルが0であるかどうかを調べる。0でなければ、ステップ231で再帰レベルをデクリメントし、プロセスはステップ226から続く。再帰レベルが0である場合、ステップ232で符号化プロセスがなされたかどうかを調べる。なされていないければプロセスはステップ223に戻る。符号化プロセスがなされていれば、ステップ233で種々の終了タスクを実行する。

【0067】(復号化) 図示された復号化のプロセスによれば、ステップ221で種々の初期化タスクを実行する。ステップ222では再帰レベルを0に初期化する。ステップ223で復号化表現を生成するためにコードを処理する。ステップ224でタグ属性が存在するかどうかを調べる。存在すれば、ステップ225で復号化表現を生成するためにタグ属性を表すコードを処理し、他のタグ属性が存在するかどうかを調べるためにステップ224に戻る。更なるタグ属性が存在しないとき、プロセスはステップ226から続く。

【0068】ステップ226でタグ内容が存在するかどうかを調べる。存在すれば、ステップ227で復号化表現を生成するためにタグ内容を表すコードを処理する。ステップ228では符号化タグ内容の中で入れ子になっているコードがあるかどうかを調べる。ない場合、他のタグ内容が存在するかどうかを調べるためにプロセスはステップ226に戻る。更なるタグ内容が存在しない場合、プロセスはステップ230から続く。符号化タグ内容の中でコードが入れ子になっている場合、ステップ229では再帰レベルをインクリメントし、プロセスはステップ223から続く。ステップ230では現在の再帰レベルが0かどうかを調べる。もし0でなければ、ステップ231にて再帰レベルをデクリメントし、プロセス

はステップ226から続く。再帰レベルが0の場合、ステップ232で復号化プロセスが終わったかどうかを調べる。もし終わっていなければ、プロセスはステップ223に戻る。復号化プロセスが終了すれば、ステップ233にて種々の終了タスクを実行する。

【図面の簡単な説明】

【図1】本発明の種々の側面が実行され得るシステムの主要な構成要素を示す図である。

【図2】ドキュメントの要素の圧縮された表現を生成するプロセス又は装置のブロック図である。

【図3】圧縮された表現からドキュメントの要素を回復するプロセス又は装置のブロック図である。

【図4】ドキュメントの要素の圧縮された表現を生成するプロセスの状態遷移図である。

【図5】圧縮された表現からドキュメントの要素を回復するプロセスの状態遷移図である。

【図6】ドキュメント情報を圧縮又は拡張するプロセスの機能的フロー図である。

【図7】マークアップ言語で表現されたシンプルなドキュメントを示す図である。

【図8】本発明による符号化プロセスによる図7のドキュメントを表す符号化された情報を示す図である。

【符号の説明】

- 11 リモート装置
- 12 ディスプレイ
- 13 ボタン
- 14、32 記憶装置
- 15、22 送信機
- 16、21 受信機
- 31 コンピュータ
- 32 記憶装置
- 40 ネットワーク
- 51、52 サーバ
- 61、63、67、71、73、77、81、110～122、140～144、147～149、151、152 パス
- 62、72 識別要素
- 64 符号化
- 66、76 組み立て
- 68、78 コードブック
- 74 復号化
- 80 処理
- 100、130 開始
- 101 符号化タグ
- 102 符号化属性名
- 103 符号化属性値
- 104 属性終了
- 105 符号化内容
- 106 内容終了
- 107、137 終了

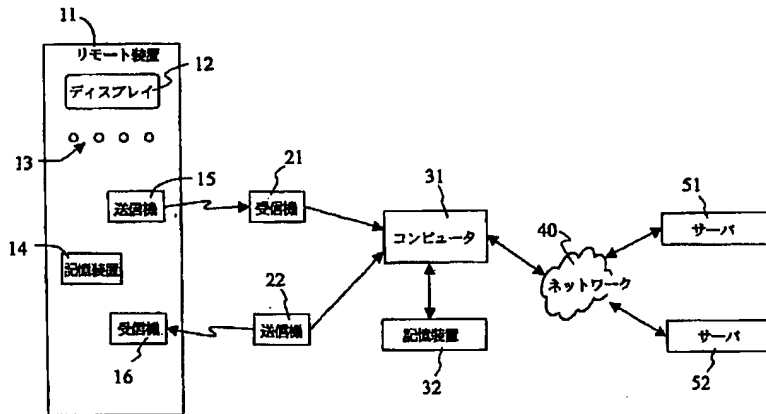
1 3 1 復号化コード

1 3 3 復号化属性値

1 3 2 復号化属性名

1 3 5 復号化内容

【図1】



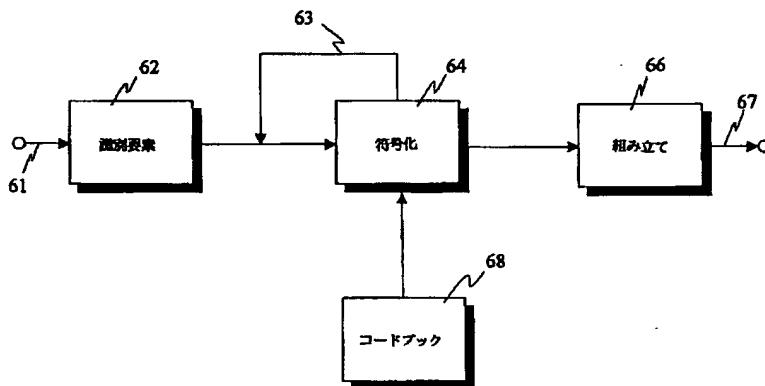
【図8】

```

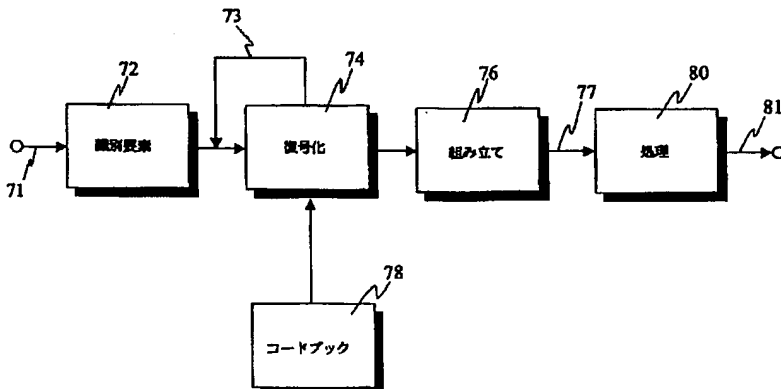
1. (HTML-C)
2. (BODY-C)
3. {STR} "The item "
4.1 {IMG-A}
4.2 {src} ("/item.gif")
4.3 {END: img-a}
5.1 {STR} "is available in "
5.2.1 {B-C}
5.2.2 {STR} "red"
5.2.3 {END: b-c}
5.3 {STR} "and "
5.4.1 {B-C}
5.4.2 {STR} "green "
5.4.3 {END: b-c}
5.5 {STR} "for a limited time. "
6.1.1 {STR} "Click"
6.1.2 {A-AC}
6.1.3 {href} ("http://a.url/info")
6.1.4 {END: a-a}
6.2.1 {STR} "hero"
6.2.2 {END: a-c}
6.3 {STR} " for more information. "
7. {END: body-c}
8. {END: html-c};

```

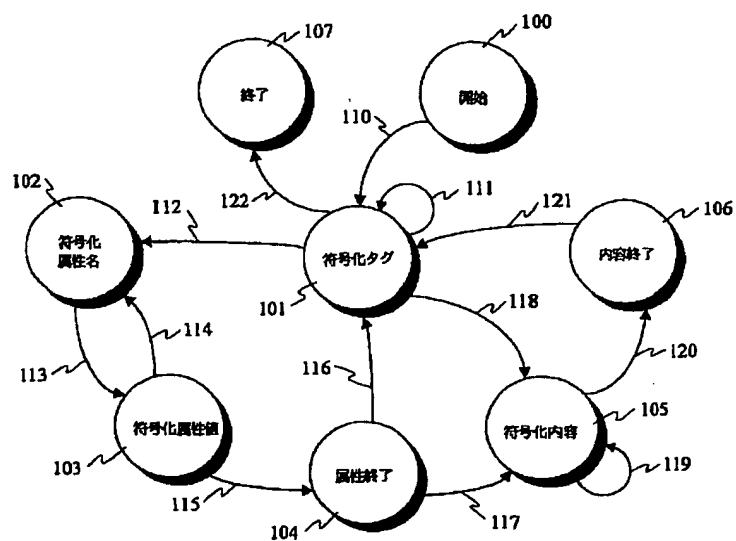
【図2】



【図3】



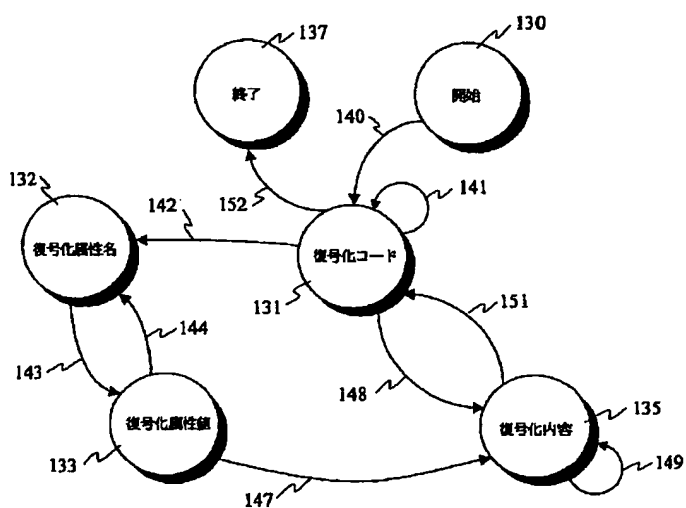
【図 4】



【図 7】

1. <HTML>
2. <BODY>
3. The item
4.
5. Is available in red and green for a limited time.
6. Click here for more information.
7. </BODY>
8. </HTML>

【図 5】



【図6】

